

SKRIPSI
PENERAPAN METODE *NAIVE BAYES CLASSIFIER*
GUNA MENGANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL X
PADA PROGRAM PENCEGAHAN *STUNTING*



SALSABILA NUR ISARA

NIM : 205610028

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2024

SKRIPSI
PENERAPAN METODE *NAIVE BAYES CLASSIFIER*
GUNA MENGANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL X
PADA PROGRAM PENCEGAHAN *STUNTING*

Diajukan sebagai salah satu syarat untuk menyelesaikan studi



Program Sarjana
Program Studi Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Disusun Oleh

SALSABILA NUR ISARA

NIM : 205610028

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA

2024

HALAMAN PERSETUJUAN

Judul : Penerapan metode *Naive Bayes Classifier*
Guna Menganalisis Sentimen Pengguna Media
Sosial X Pada Program Pencegahan *Stunting*

Nama : Salsabila Nur Isara

NIM : 205610028

Program Studi : Sistem Informasi

Program : Sarjana

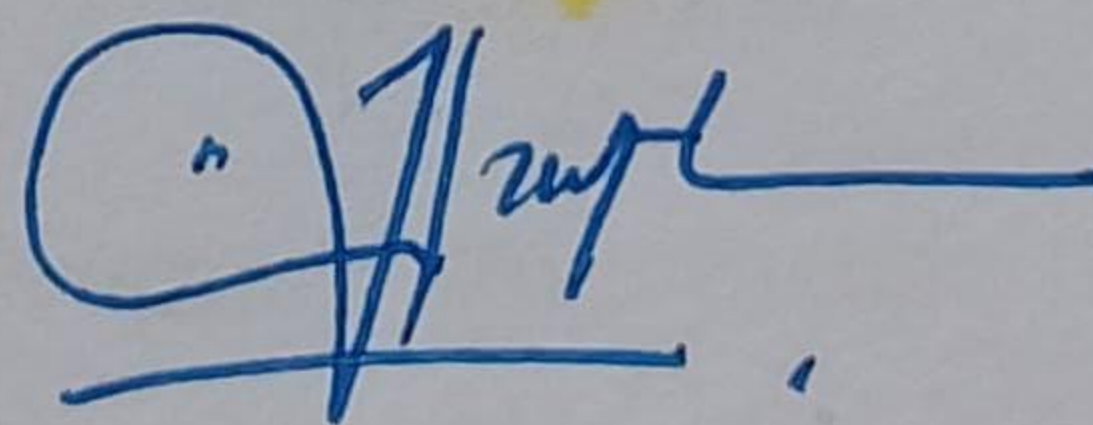
Semester : 7 (Ganjil)

Tahun Akademik : 2023/2024

Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Skripsi

Yogyakarta,.....2024

Dosen Pembimbing



Sumiyatun, S.Kom., M.Cs

NIDN : 0515048402

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN METODE *NAIVE BAYES CLASSIFIER*

GUNA MENGANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL X

PADA PROGRAM PENCEGAHAN *STUNTING*

Telah dipertahankan didepan Dewan Penguji Skripsi dan dinyatakan ditrima untuk memenuhi sebagian persyaratan guna memperoleh Gelar Sarjana Komputer

Program Studi Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta

Yogyakarta,2024

Dewan Penguji

NIDN

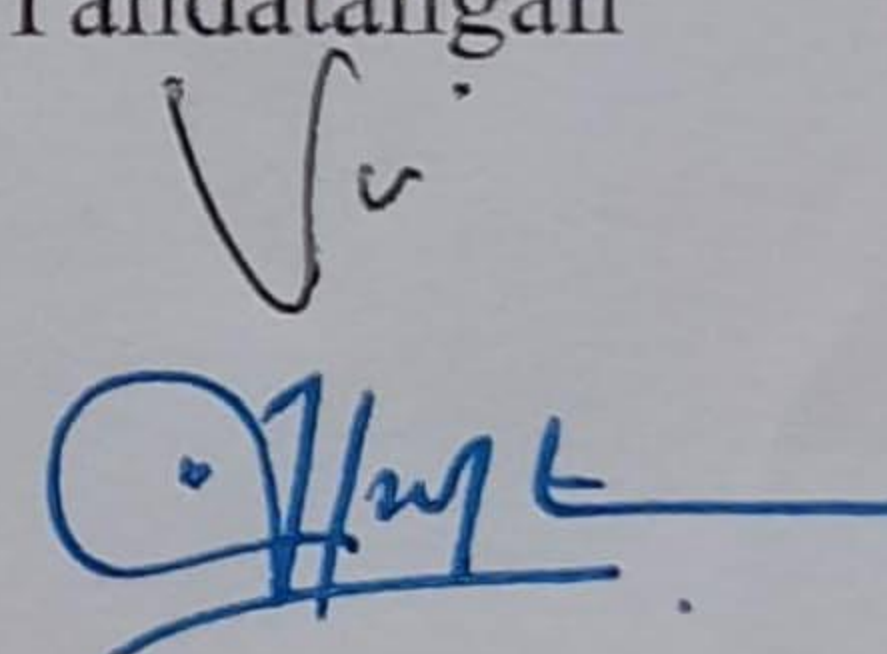
Tandatangan

1. Femi Dwi Astuti, S.Kom., M.Cs.

0516088701

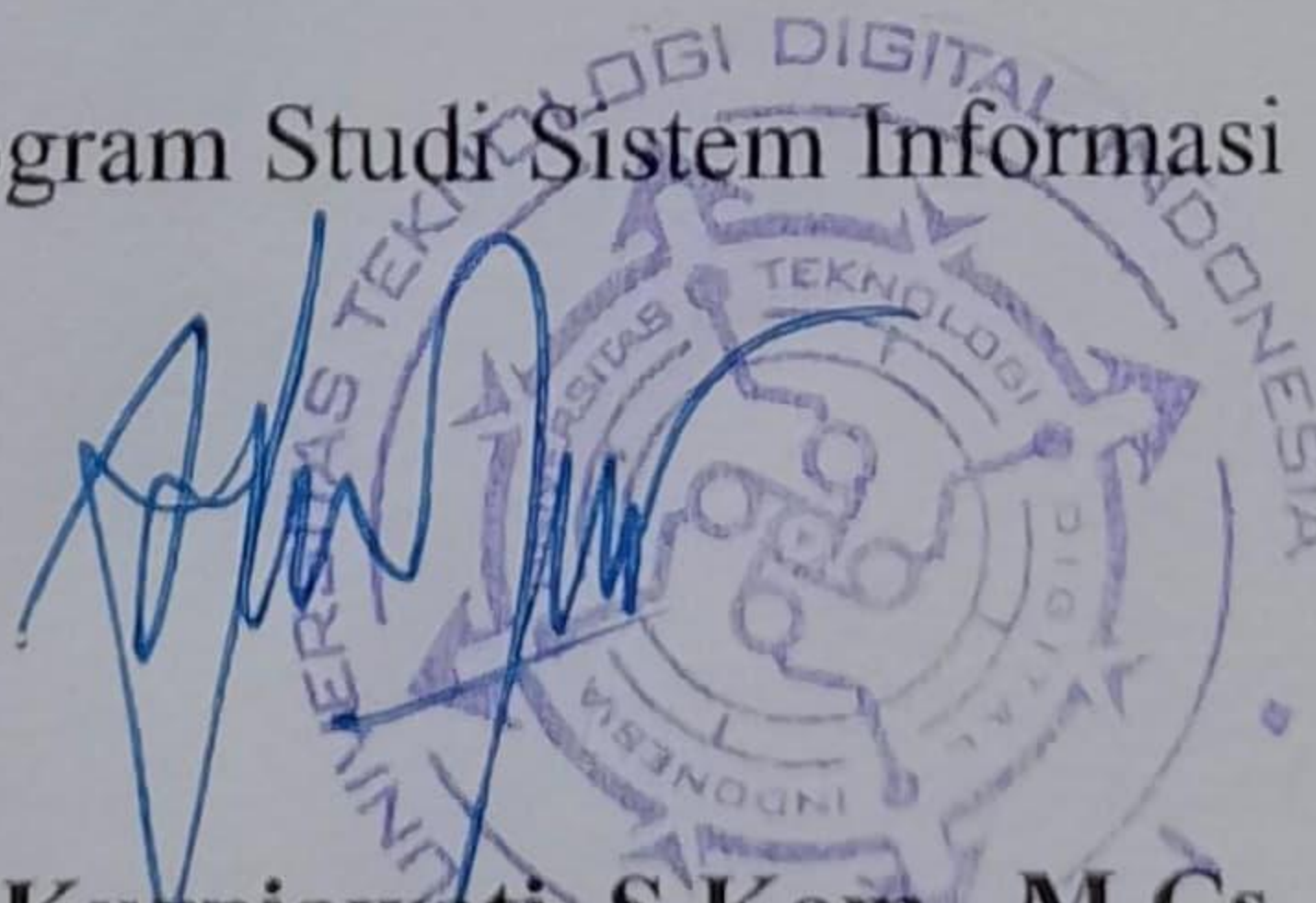
2. Sumiyatun, S.Kom., M.Cs.

0515048402



Mengetahui

Ketua Program Studi Sistem Informasi



08 MAR 2024

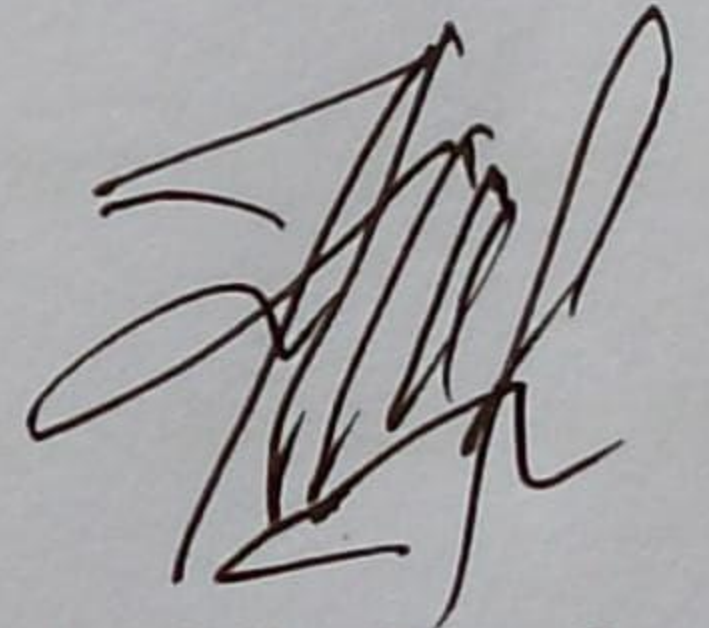
Deborah Kurniawati, S.Kom., M.Cs.

NPP : 051149

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Komputer di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 7 Maret 2024



Salsabila Nur Isara
NIM : 205610028

HALAMAN PERSEMBAHAN

Puji syukur Alhamdulillah saya panjatkan kepada Allah SWT, yang telah memberikan karunia kepada saya sehingga dapat menyelesaikan penyusunan skripsi ini yang saya persembahkan kepada :

1. Cinta pertama dan panutanku, Bapak Rohmad. Beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau sangat berperan penting dalam penyelesaian study penulis, yang selalu memberikan semangat, motivasi, serta doa yang selalu beliau berikan hingga penulis mampu menyelesaikan studynya sampai tingkat sarjana.
2. Pintu surgaku, Ibu Eka Yuliasari. Beliau juga tidak sempat merasakan pendidikan sampai bangku perkuliahan, namun beliau mampu mendidik penulis, memotivasi, memberi dukungan dan doa untuk kesuksesan anak-anaknya.
3. Nenek Samilah dan Kakek Mardi Wiyono (alm). Yang sangat ingin melihat cucu perempuannya sampai kejenjang sarjana, beliau tak hentinya mengingatkan penulis untuk selalu rajin, tekun selama menjalankan study ini, sehingga perkataan beliau yang selalu melekat dalam ingatan penulis.
4. Adikku, Varih Abitiya. Yang paling penulis sayangi walaupun sedikit menyebalkan. Terimakasih sudah ada dalam kehidupan ini dan menjadi salah satu alasan penulis semangat untuk meraih segalanya.
5. Sahabat penulis, Rati Heriyanti. yang selalu menjadi rumah ternyaman untuk berbagi keluh kesah dalam kehidupan satu sama lain

HALAMAN MOTTO

“Jangan pernah mundur dalam hal apapun, selesaikan apa yang telah dimulai”

(Penulis)

“Tidak ada kata menyerah didalam hati seorang pejuang, menang kalah biasa,
jangan menangkan kalah, mati dalam pertempuran kita siap”

(Prabowo Subianto)

KATA PENGANTAR

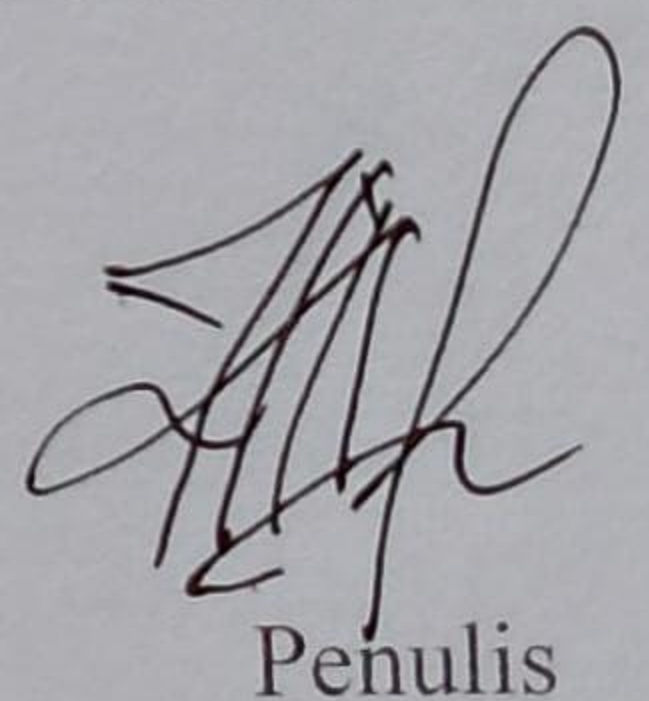
Puji syukur penulis panjatkan kehadiran Allah SWT, karena berkat rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Metode *Naive Bayes* Guna Menganalisis Sentimen Pengguna Media Sosial X Pada Program Pencegahan *Stunting* Oleh Salah Satu Pasangan Calon Presiden Dan Calon Wakil Presiden 2024” dengan baik . Penulisan skripsi ini bertujuan memenuhi salah satu syarat dalam menyelesaikan program studi S1 pada Fakultas Teknologi Informasi di Universitas Teknologi Digital Indonesia. Penulis menyadari sepenuhnya, telah banyak mendapatkan dukungan, bimbingan dan dorongan dari berbagai pihak yang telah menyumbangkan pikiran, waktu, tenaga dan sebagainya. Oleh karena itu, pada kesempatan ini dengan setulus hati penulis mengucapkan terima kasi kepada:

1. Ibu Sri Redjeki, S.Si., M.Kom., PhD, selaku Rektor Universitas Teknologi Digital Indonesia.
2. Ibu Deborah Kurniawati, S.Kom., M.Cs, selaku Ketua Program Studi Sistem Informasi.
3. Ibu Sumiyatun, S.Kom., M.Cs, selaku Dosen Pembimbing Skripsi yang telah banyak memberikan arahan bimbingan dan motivasi pada saat penulisan skripsi ini. Semoga kebaikan beliau menjadi ladang pahala Aamiin.
4. Bapak dan Ibu Dosen Sistem Informasi Universitas Teknologi Digital Indonesia Yogyakarta yang telah memberikan ilmu dari awal hingga akhir perkuliahan.

5. Kepada Bapak Rohmad dan Ibu Eka Yuliasari sebagai orang tua yang telah memberikan dukungan dan doa yang selalu menyertai sehingga penulis dapat menyelesaikan masa perkuliahan serta penyusunan skripsi ini.
6. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah banyak membantu dalam menyelesaikan skripsi ini secara langsung maupun tidak langsung. Terimakasih telah memberikan semangat dan motivasi. Terimakasih atas semuanya, terutama kebersamaan dan rasa kekeluargaan yang begitu indah selama ini.

Dalam penulisan skripsi ini, penulis sudah berusaha sebaik mungkin. Namun, masih banyak kesalahan dan kekurangan baik sengaja ataupun yang tidak disengaja. Penulis menerima kritik serta saran yang membangun sebagai perbaikan kedepannya dan semoga skripsi ini membawa manfaat bagi semua.

Yogyakarta, 7 Maret 2024



Penulis

DAFTAR ISI

SKRIPSI.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PERSEMBAHAN.....	v
HALAMAN MOTTO.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Ruang Lingkup.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penelitian.....	6
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	6
2.1 Tinjauan Pustaka.....	7
2.2 Dasar Teori.....	11
2.2.1 Analisis Sentimen.....	11
2.2.2 <i>Crawling Data</i>	12
2.2.3 <i>Tweet Harvest</i>	12
2.2.4 <i>Machine Learning</i>	12
2.2.5 Python.....	13
2.2.6 X (Twitter).....	13
2.2.7 <i>Stunting</i>	14
2.2.8 <i>Text Processing</i>	14
2.2.9 <i>TextBlob</i>	16
2.2.10 <i>Naive Bayes Classifiers</i>	18

2.2.11	<i>Confusion Matrix</i>	19
BAB III	METODE PENELITIAN.....	22
3.1	Bahan atau Data.....	22
3.2	Alat Penelitian	23
3.3	Prosedur Kerja dan Pengumpulan Data.....	23
3.3.1	Analisis.....	24
3.3.2	Pengambilan Data	25
3.3.3	<i>Pre-Processing</i> Dataset.....	26
3.3.4	Pelabelan Dataset	28
3.3.5	Klasifikasi <i>Naive Bayes</i>	30
3.3.6	Evaluasi.....	30
BAB IV	IMPLEMENTASI DAN PEMBAHASAN	31
4.1	Implementasi dan Pembahasan.....	31
4.2	Pengambilan Data.....	31
4.3	<i>Pre-Processing</i> Dataset	32
4.3.1	Cleansing.....	32
4.3.2	<i>Tokenizing</i>	33
4.3.3	Normalisasi	34
4.3.4	<i>Stopwords Removal</i>	35
4.3.5	<i>Stemming</i>	36
4.4	Pelabelan Dataset.....	37
4.5	Klasifikasi <i>Naive Bayes</i>	43
4.6	Evaluasi	45
BAB V	KESIMPULAN DAN SARAN.....	49
5.1	Kesimpulan.....	49
5.2	Saran.....	49
DAFTAR PUSTAKA	51
LAMPIRAN	53

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	24
Gambar 3. 2 Proses Pengambilan Data	25
Gambar 3. 3 Tahap PreProcessing	26
Gambar 3. 4 Tahap Pelabelan Dataset	28
Gambar 4. 1 Kode Program Crawling Data.....	31
Gambar 4. 2 Hasil Crawling Data	32
Gambar 4. 3 Kode Program Import String dan Library Regex.....	32
Gambar 4. 4 Kode Program Implementasi Cleansing.....	33
Gambar 4. 5 Hasil Cleansing	33
Gambar 4. 6 Kode Program Implementasi Tokenizing	34
Gambar 4. 7 Output Script Tokenizing	34
Gambar 4. 8 Kode Program Implementasi Normalisasi	35
Gambar 4. 9 Hasil Normalisasi	35
Gambar 4. 10 Kode Program Implementasi Stopwords Removal	36
Gambar 4. 11 Hasil Stopwords Removal.....	36
Gambar 4. 12 Kode Program Implementasi Stemming	37
Gambar 4. 13 Hasil Stemming	37
Gambar 4. 14 Kode Program Translator	38
Gambar 4. 15 Hasil Translator	38
Gambar 4. 16 Kode Program Fungsi Penentuan Sentimen dan Label.....	39
Gambar 4. 17 Kode Program Penerapan Fungsi Pelabelan dan Kolom	39
Gambar 4. 18 Hasil Implementasi Pelabelan TextBlob	40
Gambar 4. 19 Kode Program Melihat Jumlah Sentimen	41
Gambar 4. 20 Jumlah Sentimen	41
Gambar 4. 21 Kode Program Persentase Diagram Pie	41
Gambar 4. 22 Persentase Pelabelan	42
Gambar 4. 23 Kode Program Klasifikasi Naive Bayes Classifier.....	43
Gambar 4. 24 Hasil Akurasi.....	44
Gambar 4. 25 Kode Program Implementasi Tabel Confusion Matrix	45
Gambar 4. 26 Tabel Confusion Matrix	46
Gambar 4. 27 Kode Program Laporan Klasifikasi.....	47
Gambar 4. 28 Laporan Klasifikasi	48

DAFTAR TABEL

Tabel 2. 1 Perbandingan Dengan Penelitian Sebelumnya	10
Tabel 3. 1 Contoh Tweet.....	26
Tabel 3. 2 Contoh Text Cleansing.....	27
Tabel 3. 3 Contoh Teks Tokenizing	27
Tabel 3. 4 Contoh Teks Normalisasi.....	27
Tabel 3. 5 Contoh Teks Stopwords Removal.....	28
Tabel 3. 6 Contoh Teks Stemming.....	28
Tabel 3. 7 Contoh Pelabelan TextBlob	29
Tabel 3. 8 Confusion Matix	30

INTISARI

Survei asosiasi penyelenggara jasa internet Indonesia pada tahun 2023 menunjukkan bahwa jumlah pengguna internet di Indonesia mencapai 215,63 juta, meningkat 2,67% dari periode sebelumnya. Indonesia merupakan urutan ke-8 di Asia dengan penetrasi internet tertinggi. Media sosial X adalah salah satu platform yang populer di Indonesia. Analisis sentimen, menggunakan metode klasifikasi Naive Bayes, digunakan untuk memahami opini dan pandangan dalam teks, terutama dari sumber internet seperti forum, blog, media sosial, dan situs review.

Penelitian ini berfokus pada penggunaan data dari media sosial X untuk mengklasifikasikan teks bahasa Indonesia terkait kebijakan publik di bidang politik, khususnya kebijakan pencegahan stunting yang diusung oleh salah satu pasangan calon presiden dan calon wakil presiden 2024. Tweet yang dikumpulkan mulai tanggal 28 November 2023 hingga 10 Februari 2024. Tweet yang sudah terkumpul akan dilakukan pelabelan menggunakan TextBlob dan akan di klasifikasikan dengan metode Naive Bayes Classifier dengan model Confusion Matrix.

Penelitian ini menggunakan sebanyak 5.664 data, dari hasil preprocessing memperoleh sebanyak 5.630 data sampel, dari proses pelabelan diketahui bahwa terdapat 82% sentimen positif dan 18% sentimen negatif. Dari hasil klasifikasi menggunakan pembagian 80% data latih dan 20% data uji dengan 42 nilai seed acak mendapatkan akurasi sebesar 85%. Pengujian dengan model confusion matrix menghasilkan rata-rata weighted avg dari *precision* 0.84, *recall* 0.85 dan *f1-score* 0.83.

Kata Kunci: *Analysis Sentiment, Naive Bayes Classifier, Python, TextBlob.*

ABSTRACT

A survey by the Indonesian internet service providers association in 2023 shows that the number of internet users in Indonesia will reach 215.63 million, an increase of 2.67% from the previous period. Indonesia is the 8th in Asia with the highest internet penetration. Social media X is one of the popular platforms in Indonesia. Sentiment analysis, using the Naïve Bayes classification method, is used to understand opinions and views in texts, especially from internet sources such as forums, blogs, social media, and review sites.

This study focuses on using data from social media X to classify Indonesian texts related to public policy in the political field, especially stunting prevention policies carried by one of the 2024 presidential and vice presidential candidates. Tweets collected from November 28, 2023 to February 10, 2024. Tweets that have been collected will be labeled using TextBlob and will be classified with the Naïve Bayes Classifier method with the Confusion Matrix model.

This study used as many as 5,664 data, from the results of preprocessing obtained as many as 5,630 sample data, from the labeling process it was known that there were 82% positive sentiment and 18% negative sentiment. From the classification results using the division of 80% of the training data and 20% of the test data with 42 random seed values obtained an accuracy of 85%. Testing with the confusion matrix model resulted in an avg weighted average of precision 0.84, recall 0.85 and f1-score 0.83.

Keywords: Analisis Sentimen, Klasifikasi Naive Bayes, Python, TextBlob.